Don't Jump on Hype or You Will be Dumped! Using User-Generated Content to Support Financial Market Surveillance

Research in Progress

Irina Alic
University of Göttingen
Irina.alic@wiwi.uni-goettingen.de

Abstract

In this paper, an IT artifact instantiation to support decision making in the field of financial market surveillance, is presented. This artifact utilizes a qualitative multi-attribute decision model to identify situations in which prices of single stocks are affected by fraudsters who aggressively advertise the stock. A quantitative evaluation of the implemented system prototype, based on voluminous and heterogeneous data including data from social media, is provided. The empirical results indicate that that the developed IT artifact can provide support for identifying such malicious situations. Given this evidence, it can be shown that the developed solution is able to utilize massive and heterogeneous data, especially user-generated content from financial blogs and news platforms, to provide practical decision support in the field of market surveillance.

Keywords

Financial market surveillance, user generated content, market manipulation, big data.

Introduction

Financial market manipulation has received much attention from regulatory authorities, resulting in trading suspensions of those companies that may have been hijacked by fraudsters. Recent suspensions included two large groups of penny stock companies suspended by the U.S. regulatory authority on a single day (SEC 2012) before they could harm investors. All companies have been traded on over-the-counter (OTC) markets with low financial regulations. For the assessment of potentially suspicious companies, the supervisory authority used information technology to recognize thinly traded, low-capitalization companies that could be involved in fraud. The system is calibrated to detect companies with low-priced penny stocks that are traded in low volumes or not at all; such companies must also be delinquent in their public disclosures. Unfortunately, the system reveals its weak point: namely, insufficient real-time surveillance on heterogeneous data.

The pump-and-dump manipulation scheme is the most common form of information-based market manipulation in financial markets (SEC 2012). It first appears as the spreading of false positive information to market participants by fraudsters; this information, generally user-generated content on financial blogs and news platforms, is enthusiastically spread by fraudsters attempting to pump up a stock price to an artificial level. It should be noted that the stock belongs to penny stock companies in which the fraudsters already hold a significant quantity of shares. Next, uninformed market participants relying on the bogus story buy the shares, effectively forcing artificial price increases (Bouraoui 2009). Finally, fraudsters make a profit by selling the stock at the increased price level (dumping), which accordingly causes the stock price to drop.

This research addresses the question of whether the assessment of user-generated content has the potential to help regulatory authorities and financial institutions detect such situations in which companies are being manipulated. At this point, pump-and-dump abuse cases can only be studied by taking into account the recent cases that have appeared in practice. Thus, this study will present the

evaluation of a financial market surveillance system as an IT artifact instantiation that can support market surveillance in the field of information-based market manipulations such as pump-and-dump schemes.

Based on publicly-funded research project I was involved in from 2010 to 2013, the opportunity to develop an IT-based solution to tackle the mentioned case was provided. Within the research consortium consisting of research institutions and industry partners, necessary IT components have been developed in close researcher-practitioner collaboration. The problem owner (i.e., domain experts and regulatory authority) intervened according to the project needs and aligned the design principles with their surveillance issues. The development process and associated findings were reported earlier (Anonymous, 2013). During that process a surveillance system was developed as a qualitative model prototype, combining financial domain knowledge and technical expertise. At that point in time, there was no means by which to collect the data for an evaluation of the developed prototype. The developed prototype resulting from the design principles was used as a model to subsequently develop an IT artifact instantiation and implemented as a decision support system to support financial market surveillance in the field of information-based market manipulation.

With the data architecture thus in place, the aim of the present study is an evaluation of the implemented system based on a large dataset collected; for this purpose, the OTC market trading data from 1700 companies was collected during 2012 and part of 2013. The companies' data is observed and used to assess the empirical implications of market abusive behavior, with the goal of evaluating and documenting the effectiveness of the developed solution.

The objective of the present research is to demonstrate how regulatory authority can be supported by the use of both traditional data sources such as a structured time data series and additional data sources such as unstructured data collected from news platforms and blogs. The ways in which the IT artifact, utilizing user-generated content and real-time data, might ease the daily work of the financial market surveillance staff will therefore be demonstrated.

This paper is organized as follows. In the first section, related research is introduced. The second section provides an overview of the developed IT artifact. Section 3 describes the quantitative evaluation based on real data, and the final section concludes this paper.

Background

Relevant areas of study relating to this research in terms of stock fraud detection include literature on (1) stock touting impacts based on spam mails; (2) the impact of user-generated news on stock activity; and (3) financial markets services architecture. Each area will be briefly discussed in the following paragraphs.

Several studies examine of the impact of stock touting using spam emails. To generate profit, scammers spread misleading information. Authors (Frieder and Zittrain 2008) investigate market activity prior to and following a stock touting email campaign, making use of a dataset containing about 75,000 spam emails. The research reveals high market activity beginning one day before email spamming commences and continuing until the day with the most considerable number of touting mails. The authors find that volume and return respond positively to touting, whereas the returns dip significantly following the conclusion of the campaign. Hence, the study suggests two main indicators for market manipulation: abnormal price changes and volume. Research conducted by (Bouraoui 2009) demonstrates similar findings. The author provides an evolution model of volatility to assess the impact of stock spam emails based on a sample of 110 penny stock companies. The research shows abnormal returns three days after the commencement of spam emails. In both studies, the outcomes have been explained as the behavioral effect of market participants who have responded positively to the touting. Thus, these studies consider statistical approaches to explicate the influence of email stock promotion. Other studies introduce data mining techniques that help identify stock touting spam emails. Research by (Zaki et al. 2011) observes spam massages to detect highly fraudulent stock activities by utilizing data mining techniques to identify stock touting spam emails. The accuracy detection of these experiments ranges from 58% to 71%.

A considerable amount of research on the predictive power of user-generated content (such as tweets, financial forums, and blogs) on stock prices has been documented in the literature. In one instance, the research (Delort et al. 2012) introduce evidence of manipulation and examine the effect of such misuse in online financial forums. The authors show that manipulative user-generated content regarding companies

with lower prices and market capitalization positively correlates to stock returns, volatility, and volume. One recent study (Smailović et al. 2013) presents a support vector, machine-based, sentiment classifier; here, a set of about 150,000 tweets thematizing eight companies (e.g., Apple, Google, and Microsoft) served as the data basis. The authors find that positive sentiment predicts positive movement in the closing price. A study proposed by (Zhang and Skiena 2010) examines the ways in which blog and news data is reflected in trading volumes and returns. The authors demonstrate a significant positive correlation between media content and trading volume as well as stock returns.

Several commercial stock market systems for monitoring and detecting abuse in structured and unstructured data exist; some, such as the Securities Observation, News Analysis and Regulation Systems (SONAR), are presented in scientific research (Goldberg et al. 2003). SONAR, which aims to monitor the stock market, applies data mining, text mining, statistical regression, and rule-based detection to recognize both abuse patterns in structured data and unusual trading following publication of the news. A study by (Mangkorntong and Rabhi 2008) compares two different surveillance systems as event-processing systems in such areas as memory usage, scalability, and flexibility. The authors reveal the strengths and weaknesses of the two systems and suggest a generic approach that uses numerous different event-processing systems to support the detection process (Mangkorntong and Rabhi 2007).

The next section introduces the Financial Market Surveillance Decision Support System (FMS-DSS) as an IT artifact instantiation intended to support the regulatory authority in detecting malicious pump-and-dump behavior cases; with underlying software components developed by the project consortium members.

IT Artifact

Generally, decision support system configurations are built on the basis of three basic technology components: (1) data, (2) models, and (3) user interface (UI) (Turban, Sharda, Delen, 2010). Figure 1 presents the high-level components of the developed FMS-DSS. The data management component preprocesses and stores the needed data. The model component calculates the artificial changes or jumps of e.g. trading volume based on specified constraints, and processes the information and determines the alert based on predefined rules. The user interface component allows meaningful representation of the detected malice behavior.

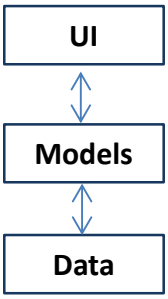

Figure 1: FMS-DSS high-level components

Initialized by the end user and during operation, the FMS-DSS searches for the appearance of user-generated content regarding the monitored company every day for a predefined time period (e.g., within the last 30 days). For this purpose, unstructured input data is continuously retrieved, preprocessed, and stored in a database. The developed software component subsequently collects other related input data. Based on problem-owner defined rules and models, the output appears as a signal, which can be either a v-high, high, medium, low, or v-low alert.

Data

Selection of potential input data categories based on pump-and-pump manipulation scheme evidence: In meetings with experts (i.e., compliance officers and the regulatory body), the specific decision problem in revealing typical factors for pump-and-dump market manipulation was explained. The three main types of information incorporated into the decision model are: the

manipulation of information concerning the company, the manipulation of the financial instrument, and news as user-generated content about the company and its financial instrument. Consequentially, the three main input data categories of market abuse suspiciousness are Company, Financial Instrument, and News:

- Company: According to the experts, there are two determiners for company suspiciousness. First, in those cases where the company is already involved in financial market manipulation, the financial authority issues litigation releases and puts the company on a blacklist, which is later refined into company, country, and industry blacklists. The second determiner of a company's reliability is its history; the manipulator often targets newer companies and companies that have gone bankrupt and have recovered again. The History attribute is thus refined into the attributes Age and Bankruptcy.
- Financial Instrument: This category is refined into the attributes Market and Trading. If a company's financial instrument is listed in a market segment with low regulatory requirements, and the company itself has low market capitalization, then this instrument is seen as an additional indicator of suspiciousness. A change in trading volume or trading behavior is also seen as suspicious.
- News: The user-generated content spread in social communities is closely analyzed by the model. The attribute News is refined into attributes Content and Sentiment; the former analyzes whether the web publication includes suspicious phrases (e.g., increase in revenue, new product development), and the latter captures the sentiment expressed within the news.

Model description

Based on further interviews, the attributes structure was transformed into a hierarchical tree, with the root node (Pump and Dump), differentiating between the pure user-generated-content data (News) and the heterogeneous data regarding the company and related financial instruments (Comp_FinInst). Hence, the tree consists of the two sub-trees: one for 'Company' and its related financial instrument 'Comp_FinInst' and the second one for 'News' (Figure 2). The proposed model aggregates the attributes into assessment of pump-and-dump market manipulation**.**

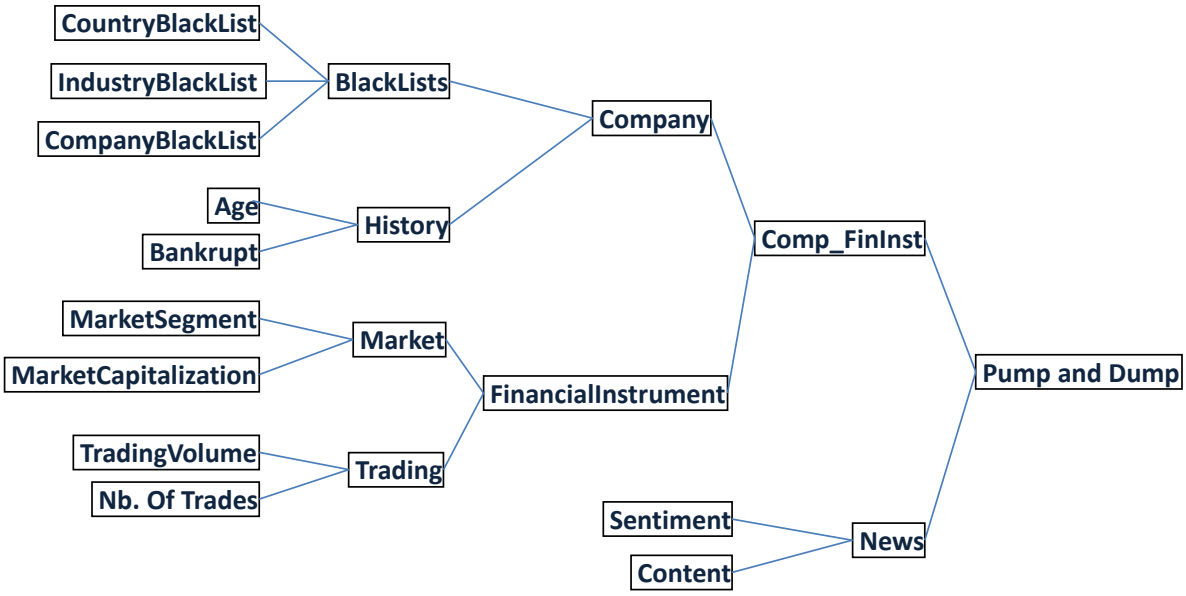

Figure 2: The hierarchical tree

Attribute scales: For each attribute, the qualitative values are scaled in the range from highly suspicious (red) to not suspicious (green), where v-low is an abbreviation for very low suspiciousness, and v-high an abbreviation for very high suspiciousness (Figure 3):

Attribute	Scale
P&D	***v-low***; low; med; high; **v-high**
Comp_FinInst	***v-low***; low; med; high; **v-high**
Company	***v-low***; low; med; high; **v-high**
BlackLists	***low***; med; **high**
CountryBlackList	***no***; **yes**
IndustryBlackList	***no***; **yes**
CompanyBlackList	***no***; **yes**
History	***low***; med; **high**
Age	***old***; med; **new**
Bankrupt	***no***; was; **is**
FinancialInstrument	***v-low***; low; med; high; **v-high**
Market	***low***; med; **high**
MarketSegment	***no***; **yes**
MarketCapitalization	***high***; med; **low**
Trading	***low***; med; **high**
TradingVolume	***low***; med; **high**
NumberOfTrades	***low***; med; **high**
News	***v-low***; low; med; high; **v-high**
Sentiment	***v-low***; low; med; high; **v-high**
Content	***v-low***; low; med; high; **v-high**

Figure 3: The attribute structure and scales

The scales for each attribute value are defined by the regulatory authority members and can be reconfigured when in use by the end user. For example, as shown in Figure 4, the default setting for the attribute 'Market Capitalization' is 'low' if the company's value is under 5 million (given currency), 'medium' if the value is between 5 and 30 million, and 'high' if the value is greater than 30 million.

Figure 4: Configuration of the attribute Market Capitalization

Manipulation scheme indicators: Within the predefined timespan the proposed calculation intends to identify abnormal changes which can be seen as indicators (Eren and Ozsoylev 2006; Goldstein and Guembel 2008; Zaki et al. 2012) for pump-and-dump abuse. Hence, the suspiciousness is assessed as follows:

Firstly, to assess recent changes in trading, long-and short-term average trading volumes are computed by taking the monthly and three-day averages of the trading volume:

$$Trading\ Volume\ Short\ Term = \frac{(\sum_{i=1}^{n} TV_i)}{n}$$

$$Trading\ Volume\ Long\ Term = \frac{(\sum_{i=1}^{m} TV_i)}{m}$$

where TV_i is the trading volume of the i-th day; n = 3 days; m = 30 days.

Secondly, to assess recent changes in number of trades, long-and short-term average trading volumes are computed by taking the monthly and three-day averages of the trading volume:

$$Number\ of\ Trades\ Short\ Term = \frac{(\sum_{i=1}^{n} NT_i)}{n}$$

$$Number\ of\ Trades\ Long\ Term = \frac{(\sum_{i=1}^{m} NT_i)}{m}$$

where NT_i is the number of trades of the i-th day; n = number of days (e.g., 3 days); m = 30 days. Additional indicators for pump-and-dump market manipulation are presented in the following Table 1.

Name	Description / Definition	Sampling
Sentiment Long-Term Period	Sentiment of News based on assessment of long-term sentiment. Based on the overall picture of the mood of the news. $SentiLong = (\sum_{I=1}^{m} Si)/m$, Where m is 4 weeks	Monthly
Sentiment Short-Term Period	Sentiment of News based on short-term sentiment (daily). Based on the overall picture of the mood of the news of one to three days. $SentiShort = (\sum_{I=1}^{n} Si)/n$, Where n is 1 day (or 3 days)	Daily
User-Generated Content Long-Term Period	Content of News based on assessment of specified terms. Based on the overall picture of the mood of the news. $ContentLong = (\sum_{I=1}^{m} Ci)/m$, Where m is 4 weeks	Monthly
User-Generated Content Short-Term Period	Content of News based on short-term specified terms (daily). Based on the overall picture of the mood of the news of one to three days. $ContentShort = (\sum_{I=1}^{m} Ci)/n$, Where n is 1 day (or 3 days)	Daily

Table 1. Calculation of average values of the input variables

Thirdly, in order to calculate jumps in e.g. price (Frieder and Zittrain 2008), the deviation of the short-term as related to the long-term average is calculated by dividing the short-term average by the long-term average and multiplying by 100, as presented in Figure 5. Three cases are assessed: when the short-term value is smaller than, greater than, or equal to the long-term value. Suspiciousness is assessed using aggregated numerical input values, which are then mapped according to qualitative scales, as defined by the problem owner as high, med and low or as v-high, high, med, low, and v-low.

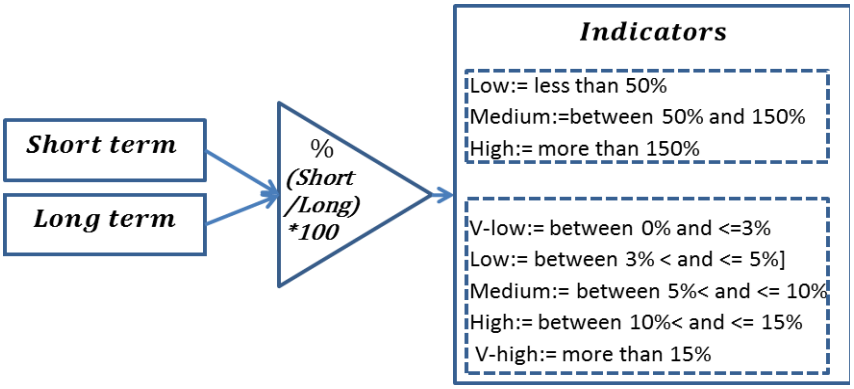

Figure 5: Calculation of artificial jumps as decision rules calibration

Accordingly, structured data (such as e.g. trading volume) and unstructured data (such as e.g. user-generated content) are thereby taken into account in order to identify abnormal changes that may be indicators of pump-and-dump market manipulations.

The recalibration of the indicator values or even the deployment of predefined default values can be adjusted by the end user, as shown in Figure 6.

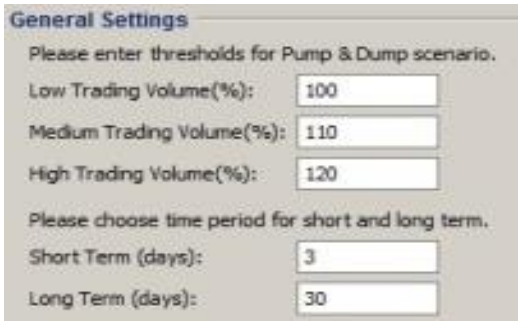

Figure 6: Decision rules calibration

Output Calculation: The final pump-and-dump alert output value (P&D), is an aggregation of the lower level attributes Comp_FinInst and News, whereas the Comp_FinInst aggregates Financial Instrument and the issuing Company. News aggregates Content and Sentiment of the News regarding the Financial Instrument and the issuing company. P&D aggregates News and Comp_FinInst and presents the final indicator which indicates whether a suspicious market situation prevails. The scales consist of five values representing the decision rules as depicted in Figure 7. For the v-high alert, three possible combinations exist; the alerts 'high' and 'med' are respectively defined by seven and eight possible combinations. The alert 'low' has six possible permutations, and the alert 'v-low' has one combination.

	Comp_FinInst	News	P&D
1	v-low	v-low	**v-low**
2	v-low	low	**low**
3	v-low	med	**low**
4	v-low	high	**med**
5	v-low	v-high	**high**
6	low	v-low	**low**
7	low	low	**low**
8	low	med	**med**
9	low	high	**med**
10	low	v-high	**high**
11	med	v-low	**low**
12	med	low	**med**
13	med	med	**med**
14	med	high	**high**
15	med	v-high	**high**
16	high	v-low	**med**
17	high	low	**med**
18	high	med	**high**
19	high	high	**high**
20	high	v-high	**v-high**
21	v-high	v-low	**low**
22	v-high	low	**med**
23	v-high	med	**high**
24	v-high	high	**v-high**
25	v-high	v-high	**v-high**

Figure 7: P&D alert output aggregation

User Interface

Figure 8 depicts a detailed view of alerts as displayed on the end-user screen. This view allows the user to drill down to a more detailed level for each alert. The related P&D output values are then stored in the database, enabling the user to post-analyze the suspect alerts.

The FMS-DSS user interface consists of a screen showing the listed 'Alerts' and 'Details' with specific information on each alert. The warnings listed in the 'Alerts' screen provide general information to the end user. Here, the company name in the list is replaced with an International Securities Identification Number (ISIN), and the exchange where the security is traded. The screen also displays the alert number under which the specific alert is stored in the database. The Date specifies the exact alert date. The user can select an alert from the list to view more detailed information.

The 'Details' screen depicts the alert selected above. Here, the display divides the information into two parts; the upper part displays unstructured data, and the lower part displays corresponding structured data. In the unstructured data area, the graph indicates the appearance of positive (green) and/or negative (red) news. The yellow line represents the difference between positive and negative news. In the alert shown in this screen shot, the yellow line appears mostly in a positive sentiment area and the light blue bars represent the sentiment amount; in this example, we note the high appearance of sentiment on the 28th and 29th March 2012.

In the structured data, we can see three colored horizontal lines showing the highest (green), lowest (red), and average (black) of daily stock prices. The dark blue vertical bars represent trading volume. The alerts, appearing as yellow dots, are evident on the 29th and 30th of March and the 3rd of April.

On the right side of the screen comprises some of the phrases collected from social-net news that have been used in the calculation of sentiment values; this feature allows the end user to observe the corresponding text and follow it to its source and author.

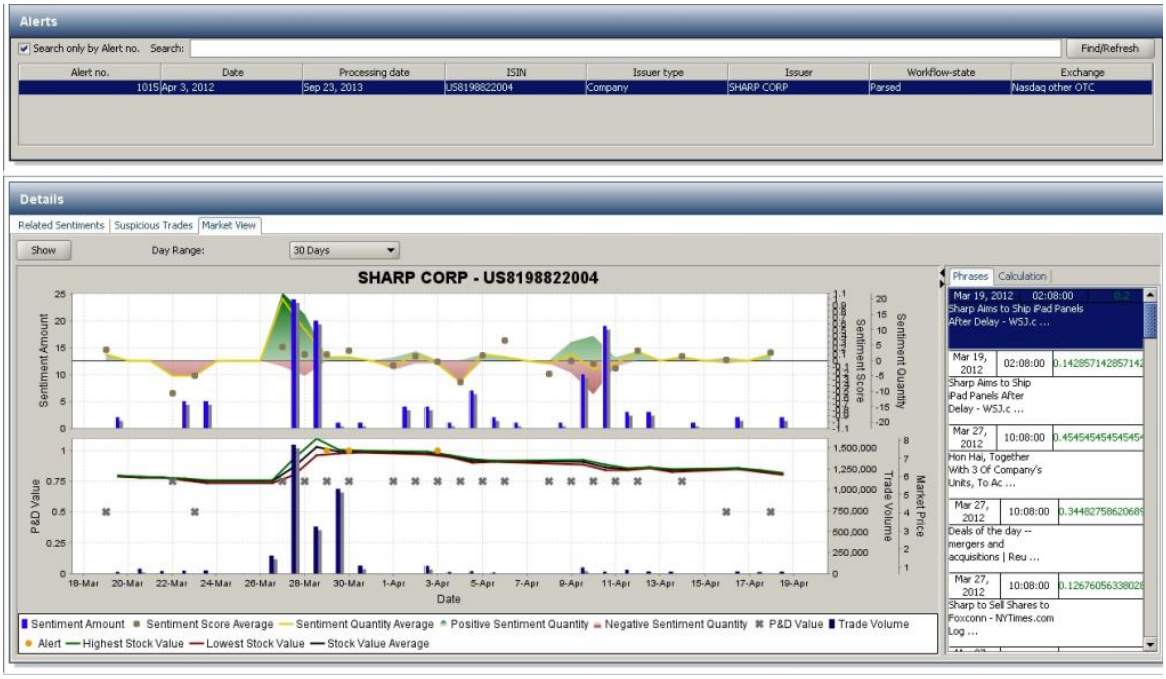

Figure 8: FMS-DSS user interface as detail view of an alert

Also of the right side of the screen, the Calculation tab contains further information on the calculation of the P&D output alert, as illustrated in Figure 9; the example data shown in Figure 9 reveals a stock market price change of 42% in addition to further relevant attribute values with their corresponding scales. For the user's convenience, a history of alerts and the corresponding database ID of the specific company are also listed. The optional 'Threshold' value, which specifies the rate of price change, can be edited by the end user. In this example, an end user has configured the system to only select stocks where the daily price change is higher than 20%.

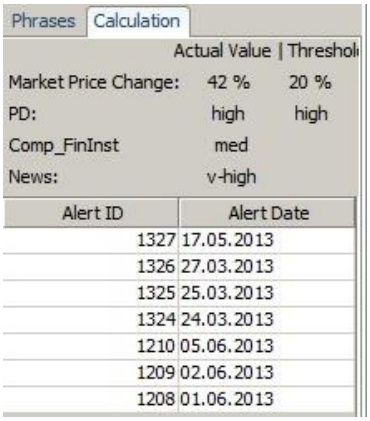

Figure 9: An alert on calculation level

The end user is thus provided with a range of information to use in the task of financial market monitoring.

The next section introduces the evaluation of output data of the FMS-DSS intended to support the regulatory authority in detective malice pump-and-dump behavior cases.

IT Artifact Evaluation

This section presents an empirical study of the FMS-DSS described in the previous section. In particular, the P&D output data is used for the performance assessment. This section is organized as follows: first, the data required for the evaluation will be described; then, after presenting descriptive statistics of this data, the relationship between the P&D output values and actual observed stock price changes is analyzed.

Dataset Description

There are two different sources of unstructured data considered in this dataset: user-generated content gathered from a variety of financial blogs, and news platforms. The unstructured input data is preprocessed and stored in a database, which is then assessed by the other system components as described previously; the data publicized at regulatory authority pages is likewise stored as input data in the database. Structured financial data is automatically downloaded by the system from a data vendor (which was involved as one industry partner in the research project), while regulatory data regarding the blacklist is stored as a list in a database and can be modified by the end user. Overall, three different data sources are considered: user-generated content, regulatory authority data, and stock price data for the period from 01.01.2012 to 03.09.2013. As a result, the P&D output sample generated by the FMS-DSS consists of 1700 OTC stock trades, with the corresponding structured and unstructured data containing 118096 entries.

Descriptive Analysis

The P&D output data is clustered within the five alert groups: very high, high, medium, low, and very low suspiciousness. The evaluations of suspiciousness in the given dataset of 118096 entries are presented in Table 2.

Alert	Count	Percentage
v-high	982	0.83%
High	22215	18.81%
Med	92062	77.96%
Low	2780	2.35%
v-low	57	0.05%

Table 2. Examination of the highest and lowest daily price changes

Statistical analysis reveals prevalent medium- and high-value alerts. This may be explained by the sensitivity of the model. Certainly the model is developed as a qualitative model where pre-defined default values might be too receptive. Allowing the end user to configure their own aggregation rules (as presented in Figure 7) and set their own thresholds (as presented in Figure 9) would therefore allow for improvement in alert sensitivity, to the effect that false positive alerts can be reduced.

The output data presented above already incorporates abnormal volume changes as a market abuse indicator. In the following section, I will examine whether the P&D output is related to the stock price changes.

Manipulation Examination

According to the domain experts involved in the research project, v-high and high alerts are the most relevant situations that demand further investigation of individual cases. On that account, the evaluation proceeds by grouping v-high and high alerts into a highly suspicious class (v-h, h), and the other alerts into a lower suspiciousness class (m, l, v-l).

Significant price changes as an indicator (Hanke and Hauser 2008; Zaki et al. 2011) of stock manipulation is applied in order to determine if the developed surveillance system can provide support to detect potentially suspicious pump-and-dump abuse cases. In order to evaluate the system's capability to detect such relevant cases, actual price changes in situations that were assessed by the system as highly suspicious (v-h, h) are compared with those that were assessed as less suspicious (m, l, v-l). In doing so, for each v-high, high, medium, low and v-low P&D output value, the changes between the highest (p_H) and the lowest (p_L) daily stock price at the alert date are calculated. To assess the changed market price value for the lowest and highest daily prices, the following measure is calculated:

$$f = \frac{p_H - p_L}{p_H} * 100$$

A hypothesis is thus formulated to verify whether actually observed price changes related to the (v-h, h) class are significantly higher than the actual price changes related to the (m, l, v-l) class:

$$H_0: \Delta p_{(v-h,h)} \leq \Delta p_{(m,l,v-l)}$$
$$H_1: \Delta p_{(v-h,h)} > \Delta p_{(m,l,v-l)}$$

In doing so, and given suitably sized samples ($N_{v\text{-}h,\ h}$ =23197; $N_{m,l,v.l}$=94899) a *t*-test is applied to test the formulated hypothesis. The null hypothesis can be rejected at the 1% level. The results are presented in Table 3.

	N=118096	Mean	Std. dev.	Median
(v-h, h)	N=23197	1.24	3.65	0.41
(m, l, v-l)	N=94899	0.4	1.46	0
p-value	<0.01			

t-value	34.31

Table 3. Evaluation results

Undoubtedly, it appears that focusing on the group of v-high and high alerts is advantageous, as the t-value is highly significant. Hence, it can be stated that suspicious abuse cases where fraudsters first buy shares and then attempt to manipulate their price by spreading extremely euphoric content on social media, can be detected by the implemented FMS-DSS. This statistical evaluation encourages in such a way that the default values as defined by the practitioners and regulatory authority might be seen as the best default values.

A potential shortcoming from the end-user perspective might be the huge number of potentially suspicious daily alerts (55 in averages) the system provides. Therefore, in order to reduce the overwhelming number of positive alerts to capture only the most probable suspect patterns, further refinements might be undertaken by the end user, such as a threshold routine that calculates the changes between the highest and the lowest daily stock price. For every price change greater than or equal to a specified percentage (in our previous example 20%), the routine will trigger an alert. An examination of such price changes results in a reduced list of 268 alerts, in other words, about one alert every two days. Thus, the procedure might reduce the noise of less relevant alerts and pinpoint the most suspicious manipulation cases. It appears that filtering based on price changes significantly increases the appearance of v-high values from 0.83% to 23.88% and of high values from 18.81% to 34.33%. The appearance of a medium alert, however, falls significantly from 77.96% to 41.04%. Hence, the final procedure will filter the most promising suspect alerts and help reduce their daily appearance to the lowest reasonable number.

Conclusion

This work has documented the effectiveness of the FMS-DSS and the importance of a market surveillance solution in improving the detection of information-based market manipulation. It has shown how the regulatory authority can be supported by the use of a market surveillance service. The evaluation results show that user-generated content can be processed in real time using the developed solution. The P&D alert results show that the solution is able to handle large-scale data and provide timely daily alerts by distinguishing suspect and non-suspect values. Most notably, this is the first study that investigates the effectiveness of market surveillance decision support that considers three different data sources: user-generated content, regulatory authority data, and time series data from a data vendor. Our results provide convincing evidence for a long-term analysis (of approximately two years) of real data and suggest that the developed solution may be effective in detecting real abusive cases of pump-and-dump market manipulation. However, some limitations are worth noting; even if the research is supported statistically, the solution has not yet been evaluated in the real-world conditions of a compliance office. Future research should therefore include a subsequent effort in order to evaluate the acceptance and use of the running solution.

REFERENCES

Anonymous, 2013

Bouraoui, T. 2009. "Stock spams: An empirical study on penny stock market," *International Review of Business Research Papers* , pp. 1–14.

Delort, J.-Y., Arunasalam, B., Leung, H., and Milosavljevic, M. 2012. "The impact of manipulation in internet stock message boards," *International Journal of Banking and Finance* (8:4).

Eren, N., and Ozsoylev, H. N. 2006. "Hype and dump manipulation," in *EFA 2007 Ljubljana Meetings Paper, AFA*.

Frieder, L., and Zittrain, J. 2008. "Spam works: Evidence from stock touts and corresponding market activity," *Hastings Communications and Entertainment Law Journal* (479).

Goldberg, H., Kirkland, J., Lee, D., Shyr, P., and Thakker, D. 2003. "The NASD Securities Observation, New Analysis and Regulation System (SONAR)," *IAAI* , pp. 11–18.

Goldstein, I., and Guembel, A. 2008. "Manipulation and the allocational role of prices," *Review of Economic Studies* (75:1)Wiley Online Library, pp. 133–164.

Hanke, M., and Hauser, F. 2008. "On the effects of stock spam e-mails," *Journal of Financial Markets* (11:1), pp. 57–83.

Mangkorntong, P., and Rabhi, F. 2007. "A High-Level Approach for Defining & Composing Event Patterns and Its Application to E-Markets," in *The Second International Workshop on Event-driven Architecture, Processing and Systems (EDA-PS 2007) at the 33rd International Conference on Very Large Data Bases (VLDB 2007)*, pp. 1–4.

Mangkorntong, P., and Rabhi, F. a. 2008. "A Domain-Driven Approach for Detecting Event Patterns in E-Markets," *World Wide Web* (12:1), pp. 69–86.

SEC, G. 2012. "SEC Suspends Trading in Common Stock of Three Hundred Seventy- Nine Compnies Quoted on OTC May 14 , 2012," *SEC*.

Smailović, J., Grčar, M., Lavrač, N., and Žnidaršič, M. 2013. "Predictive Sentiment Analysis of Tweets: A Stock Market Application," in *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 77–88.

Turban, R. Sharda, and D. Delen. *Decision support and business intelligence systems*, (9th ed.) Prentice Hall, New Jersey (2010).

Zaki, M., Diaz, D., and Theodoulidis, B. 2012. "Financial Market Service Architectures: A 'Pump and Dump' Case Study," *2012 Annual SRII Global Conference* Ieee, pp. 554–563.

Zaki, M., Theodoulidis, B., and Solis, D. D. 2011. "A Data Mining Approach for the Analysis of 'Stock-Touting' Spam Emails," *Journal of Manufacturing Technology Management* (22:6), pp. 70–79.

Zhang, W., and Skiena, S. 2010. "Trading Strategies to Exploit Blog and News Sentiment.," in *4th Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*.